Stochastic approaches for modelling in vivo reactions

T. E. Turner [a], S. Schnell [b,c,*] and K. Burrage [d].

[a] *Oxford Centre for Industrial and Applied Mathematics, Mathematical Institute, 24-29 St. Giles', Oxford OX1 3LB, UK*

[b] *Centre for Mathematical Biology, Mathematical Institute, 24–29 St Giles', Oxford OX1 3LB, UK*

[c] *Christ Church, Oxford OX1 1DP, UK*

[d] *Department of Mathematics, Advanced Computational Modelling Centre, University of Queensland, Brisbane QLD4072, Australia*

Abstract

In recent years, stochastic modelling has emerged as a physically more realistic alternative for modelling in vivo reactions. There are numerous stochastic approaches available in the literature; most of these assume that observed random fluctuations are a consequence of the small number of reacting molecules. We review some important developments of the stochastic approach and consider its suitability for modelling intracellular reactions. We then describe recent efforts to include the fluctuation effects caused by the structural organisation of the cytoplasm and the limited diffusion of molecules due to macromolecular crowding.

Key words: intracellular reactions; stochastic simulation algorithm; τ-leap method; quasi-steady-state approximation; fractal-like kinetics.

* Corresponding author. Telephone: +44-(0)1865-286855. Fax: +44-(0)1865-270515.
Email addresses: tom.turner@cantab.net (T. E. Turner), schnell@maths.ox.ac.uk (S. Schnell), kb@maths.uq.edu.au (K. Burrage).

1 Introduction

Dramatic advances in genetic and molecular biology have brought about an unprecedented flood of genomic data. Analysis of this data – to understand how genes and proteins work collectively – has led to a significant increase in the use of computers both for modelling and data interpretation. This is one of the challenges of modern science. The current interest in computational cell biology reflects the widespread belief that the complexity and sophistication of computers and programming could potentially match the complexity of living cells. The aim of computational biology is to produce sophisticated computer simulations against which biological phenomena, data or patterns are compared. Unfortunately, no consensus presently exists as to the best methodologies for performing these tasks. This is particularly true for the computational modelling of complex biochemical reactions and gene networks in cellular media.

The modelling of chemical reactions using deterministic rate laws has proven extremely successful in both chemistry (Epstein and Pojman, 1998) and biochemistry (Heinrich and Schuster, 1996) for many years. This deterministic approach has at its core the law of mass action, an empirical law giving a simple relation between reaction rates and molecular component concentrations. Given knowledge of initial molecular concentrations, the law of mass action provides a complete picture of the component concentrations at all future time points (Espenson, 1995).

The law of mass action considers chemical reactions to be macroscopic under convective or diffusive stirring, continuous and deterministic (Cox, 1994). These are evidently simplifications, as it is well understood that chemical reactions involve discrete, random collisions between individual molecules. As we consider smaller and smaller systems, the validity of a continuous approach becomes ever more tenuous. As such, the adequacy of the law of mass action has been questioned for describing intracellular reactions (Clegg, 1984; Halling, 1989; Kuthan, 2001). Arguments for the application of stochastic models for chemical reactions come from at least three directions, since the models (a) take into consideration the discrete character of the quantity of components and the inherently random character of the phenomena; (b) are in accordance (more or less) with the theories of thermodynamics and stochastic processes; and (c) are appropriate to describe "small systems" and instability phenomena.

More than 150 years ago the Scottish botanist Robert Brown discovered the existence of fluctuations whilst studying microscopic living phenomena (for the early history

of Brownian motion, Kerker, 1974). This in itself has significant implications for biochemistry where we often wish to model reaction rates within individual cells where the volume of the system is small and the molecular populations often too low for the system to be considered *macroscopic*. At the molecular level, random fluctuations are inevitable, with their effect being most significant when molecules are at low numbers in the biochemical system. This typically occurs in the regulation of gene expression where transcription factors interact with DNA binding sites in the gene's regulatory sequences. Indeed, these intrinsic fluctuations have recently been measured using fluorescent probes (see, for example, Elowitz et al., 2002; Blake et al., 2003). Additionally, it has been proven that low copy numbers of expressed RNAs can be significant for the regulation of downstream pathways (McAdams and Arkin, 1997). Thus, there are evidently a number of important biological environments where only small numbers of molecules are present in the reaction volume, for which, it is argued, stochastic modelling approaches are required (Morton-Firth and Bray, 1998).

There is also growing evidence of the importance for reaction kinetics of the structural organisation of the intracellular environment, which is far from the homogeneous, well mixed solution typical of in vitro experiments (see Schnell and Turner, 2004, and references therein). Cellular environments are highly compartmented and structured throughout the reaction volume. A high degree of molecular crowding as well as the presence of endogenous obstacles in cellular media have important consequences in the thermodynamics of the cell (Minton, 1993, 1998) and strongly affect diffusion processes (Luby-Phelps et al., 1987). The viscosity of the mitochondrion is 25–37 times higher than that of a typical in vitro experimental buffer (Scalettar et al., 1991). Diffusion of macromolecules in the cytoplasm can be 5–20 times lower than in saline solutions (Verkman, 2002). Furthermore, many reactions occur on two-dimensional membranes or one-dimensional channels (Clegg, 1984; Srere et al., 1989). These structural considerations mean we must be careful when considering how *well mixed* a chemical system is.

The stochastic approach introduced above uses the inherent random nature of microscopic molecular collisions to build a probabilistic model of the reaction kinetics (Qian and Elson, 2002). This approach is thus inherently suited to the small, heterogenous environments typical of in vivo conditions (Kuthan, 2001). However, one major problem with stochastic methods is that they are difficult to implement analytically and researchers are reduced to numerical studies.

In this work, we review some important developments of the stochastic approach

and consider its suitability for modelling in vivo reactions. Firstly, we examine how Gillespie (1977) has used the stochastic formalism to develop an algorithm for simulating reaction dynamics, and illustrate by means of numerical simulations how the stochastic and deterministic approaches compare. We discuss some further streamlining of the algorithm by Gibson and Bruck (2000), Gillespie (2001), Burrage and Tian (2003), Rathinam et al. (2003) and Tian and Burrage (2004a) leading to greater computational efficiency. We then consider how Rao and Arkin (2003) have incorporated the quasi-steady state assumption – an approximation derived from the deterministic approach – into the stochastic method, and the computational savings achieved. We highlight the failure of the considered stochastic approaches to incorporate non-homogeneities typical of in vivo conditions into the models and present an alternative two-dimensional Monte-Carlo approach by Berry (2002) and Schnell and Turner (2004). Finally we discuss the implications of Schnell and Turner's results – in particular with relation to Kopelman's formulation of fractal-like kinetics – to our understanding of in vivo biochemical kinetics.

2 The deterministic and stochastic approaches

There is now considerable evidence from both theoretical and experimental perspectives of the role of noise in biochemical pathways. Fedoroff and Fontana (2002) remark that "stochasticity is evident in all biological processes. The proliferation of both noise and noise reduction is a hallmark of organismal evolution". However, a natural question to ask is: What is the nature of this stochasticity? Hume (2000) notes that "transcription in higher eukaryotes occurs with a relatively low frequency in biological time and is regulated in a probabilistic manner". Sano et al. (2001) also remark that "initiation of gene transcription is a discrete process in which individual protein-coding genes in an off state can be stochastically switched on, resulting in sporadic pulses of mRNA production". This is the dichotomy that we must resolve – proteins are discrete objects, yet their effects are often modelled (as ordinary differential equations) in terms of concentrations.

Recently, Crampin and Schnell (2004) pointed out that "biological systems are characterised by their regulatory and adaptive properties, from homeostatic mechanisms which maintain constant output levels to switching between alternative substrates or developmental pathways. Regulatory mechanisms including thresholds, allosteric interactions and feedback in gene transcription networks, metabolic pathways, signal transduction and intercellular interactions are defining biological characteristics—

almost everything that happens in life boils down to enzyme-catalysed reactions". This leads us to the modelling process of how to represent in vivo enzymic reactions mathematically. There are many approaches; these include:

- Directed graphs in which molecules are vertices and the reactions are the edges;
- Bayesian networks in which the vertices correspond to random variables that describe, for example, a gene expression while the network defines a joint probability density function;
- Boolean networks in which a biological object is either in an on or off state;
- Ordinary Differential Equations (ODEs) in which chemical kinetics rate equations are used to represent protein concentrations;
- Partial Differential Equations (PDEs) in which the spatial structure of cells are taken into account; and finally
- Stochastic Differential Equations (SDEs) in which we have to resolve the issue of whether we work with concentrations or with individual molecules through continuous or discrete models.

As the previous discussions would suggest, we can consider three different types of modelling regimes for understanding biochemical pathways and networks. These are the discrete and stochastic, the continuous and stochastic and the continuous and deterministic regimes and reflect the nature of the considered reactions and number of molecules present in the system.

2.1 Deterministic: The law of mass action

The fundamental empirical law governing reaction rates in biochemistry is the *law of mass action*. This states that for a reaction in a homogeneous, free medium, the reaction rate will be proportional to the concentrations of the individual reactants involved. For example, given the simple Michaelis–Menten reaction

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \xrightarrow{k_2} E + P \tag{1}$$

the rate of production of complex C would be

$$\frac{\mathrm{d}C_+}{\mathrm{d}t} = k_1 SE$$

and the rate of destruction of C would be

$$\frac{\mathrm{d}C_-}{\mathrm{d}t} = k_{-1}C + k_2 C \ .$$

Combining these terms gives an expression for the rate of change of concentration of C

$$\frac{\mathrm{d}C}{\mathrm{d}t} = \frac{\mathrm{d}C_+}{\mathrm{d}t} + \frac{\mathrm{d}C_-}{\mathrm{d}t} = k_1 SE - (k_{-1} + k_2)C \ . \tag{2}$$

Using this law, similar expressions for the rate of change of concentration of each of the molecules can be found. Hence, we can express any chemical system as a collection of coupled non-linear first order differential equations. Apart from the most simple cases these do not in general have an analytical solution (Schnell and Mendoza, 1997). However, it is straightforward enough to numerically integrate them to find an approximation of the reaction dynamics of the system.

2.2 Stochastic: The chemical master equation

Whereas the deterministic approach outlined above is essentially an empirical law, derived from in vitro experiments, the stochastic approach is far more physically rigorous. The stochastic treatment of chemical reactions was initiated by Kramers (1940). Fundamental to the principal of stochastic modelling is the idea that molecular reactions are essentially random processes; it is impossible to say with complete certainty the time at which the next reaction within a volume will occur. In macroscopic systems, with a large number of interacting molecules, the randomness of this behaviour averages out so that the overall macroscopic state of the system becomes highly predictable. It is this property of large scale random systems that enables a deterministic approach to be adopted; however, the validity of this assumption becomes strained in in vivo conditions as we examine small-scale cellular reaction environments with limited reactant populations.

Bartholomay (1957) was one of the first biochemists to examine enzyme-catalysed reactions within the framework of statistical kinetics. Over the subsequent 20 years the framework led to the stochastic analysis of a variety of simple reaction mechanisms including the Michaelis–Menten mechanism (Bartholomay, 1962a,b; Jachimowski et al., 1964; Darvey and Staff, 1967; Staff, 1970; Arányi and Tóhn, 1977). As explicitly derived by Gillespie (1992b), the stochastic model uses basic Newtonian physics and thermodynamics to arrive at a form often termed the *propensity function* that gives the probability a_μ of reaction μ occurring in time interval $(t, t + \mathrm{d}t)$

$$a_\mu \mathrm{d}t = h_\mu c_\mu \mathrm{d}t \ , \tag{3}$$

where the M reaction mechanisms are given an arbitrary index μ $(1 \leq \mu \leq M)$ and h_μ denotes the number of possible combinations of reactant molecules involved

in reaction μ. For example, if reaction l involves two species S_1 and S_2 with X_i molecules of species S_i, we have $h_l = X_1 X_2$.

The rate constant c_μ is dependent on the radii of the molecules involved in the reaction, and their average relative velocities – a property that is itself a direct function of the temperature of the system and the individual molecular masses (Gillespie, 1977). These quantities are basic chemical properties which for most systems are either well known or easily measurable. Thus, for a given chemical system, the propensity functions, a_μ can be easily determined. Indeed, their form as described above, constitute the *fundamental hypothesis* of the stochastic formulation of chemical kinetics – valid for any chemical system that is kept *"well mixed"* either by direct stirring or by requiring that non-reactive molecular collisions occur far more frequently than reactive molecular collisions (Gillespie, 1976).

The stochastic formulation proceeds by considering the *grand probability function* $P(\mathbf{X}; t) \equiv$ probability that there will be present in V at time t, X_i of species S_i, where $\mathbf{X} \equiv (X_1, X_2, ..., X_N)$ is a vector of molecular species populations. Evidently, knowledge of this function provides a complete understanding of the probability distribution of all possible states at all times.

By considering a discrete infinitesimal time interval $(t, t + \mathrm{d}t)$ in which either 0 or 1 reactions occur [1] we see that there exist only $M + 1$ distinct configurations at time t that can lead to the state $\mathbf{X}$ at time $t + \mathrm{d}t$ and as such, we can write our grand probability function at time $t + \mathrm{d}t$ as a function of all possible precursor states at time t

$$\begin{aligned} P(\mathbf{X}; t + \mathrm{d}t) = {} & P(\mathbf{X}; t) P(\text{no state change over } \mathrm{d}t) \\ & + \textstyle\sum_{\mu=1}^{M} P(\mathbf{X} - \mathbf{v}_\mu; t) P(\text{state change to } \mathbf{X} \text{ over } \mathrm{d}t) \ , \end{aligned}$$

where $\mathbf{v}_\mu$ is a stoichiometric vector defining the result of reaction μ on state vector $\mathbf{X}$, i.e. $\mathbf{X} \to \mathbf{X} + \mathbf{v}_\mu$ after an occurrence of reaction μ. It is straightforward to show that

- P(no state change over dt) $= 1 - \sum_{\mu=1}^{M} a_\mu(\mathbf{X})\mathrm{d}t$.
- P(state change to $\mathbf{X}$ over dt) $= \sum_{\mu=1}^{M} P(\mathbf{X} - \mathbf{v}_\mu; t) a_\mu(\mathbf{X} - \mathbf{v}_\mu)\mathrm{d}t$.

[1] The probability of more than one reaction occurring in time interval $(t, t + \mathrm{d}t)$ is $o(\mathrm{d}t)$ and hence vanishes in the limit $dt \to 0$

If we then note that

$$\lim_{\mathrm{d}t\to 0} \frac{P(\mathbf{X}; t+\mathrm{d}t) - P(\mathbf{X}; t)}{\mathrm{d}t} = \frac{\partial P(\mathbf{X}; t)}{\partial t}$$

we arrive at the *chemical master equation* that describes the stochastic dynamics of the system

$$\frac{\partial P(\mathbf{X}; t)}{\partial t} = \sum_{\mu=1}^{M} a_\mu(\mathbf{X} - \mathbf{v}_\mu) P(\mathbf{X} - \mathbf{v}_\mu; t) - a_\mu(\mathbf{X}) P(\mathbf{X}; t) \ . \tag{4}$$

3 Stochastic simulation algorithms

Essentially the characterisations of the three modelling regimes — the discrete and stochastic, the continuous and stochastic and the continuous and deterministic — depend on the nature of the reactions and the number of molecules in the system being studied. One key simulation technique is the stochastic simulation approach to chemical reactions developed by Gillespie (1977) through the stochastic simulation algorithm (SSA). This is essentially an exact procedure for numerically simulating the time evolution of a well-stirred chemically reacting system by taking proper account of the randomness inherent in such a system. It is rigorously based on the same microphysical premise that underlies the chemical master equation described above (Gillespie, 1992a) and gives a more realistic representation of a system's evolution than the deterministic reaction rate equation (RRE) represented mathematically by ODE. In particular, the RRE is entirely inappropriate if the molecular population of some critical reactant species is so small that microscopic fluctuations can produce macroscopic effects. This is especially true for the enzymatic reactions in living cells (Kuthan, 2001). As with the chemical master equation, the SSA converges, in the limit of large numbers of reactants, to the same solution as the law of mass action.

The algorithm takes time steps of variable length, based on the rate constants and population size of each chemical species. The probability of one reaction occurring relative to another is dictated by their relative propensity functions. According to the correct probability distribution derived from the statistical thermodynamics theory, a random variable is then used to choose which reaction will occur, and another random variable determines how long the step will last. The chemical populations are altered according to the stoichiometry of the reaction and the process is repeated. In recent years, the SSA has been successfully applied in a number of settings including λ-phage (Arkin et al., 1998), and circadian rhythms (Elowitz and Leibler, 2000; Gonze et al., 2002). The cost of this detailed stochastic simulation algorithm is the

likely large amounts of computing time. The key issue is that the time step for the next reaction can be very small indeed if we are to guarantee that only one reaction can take place in a given time interval.

An alternative approach to the SSA is via the **StochSim** package developed initially by Morton-Firth (1998) [now Carl Firth] as part of a study of bacterial chemotaxis. The aim was to develop a realistic way of representing the stochastic features of this signalling pathway and to handle the large numbers of individual reactions encountered (Firth and Bray, 2000). Molecules or molecular complexes are represented as individual software objects. Reactions between molecules occur stochastically, according to probabilities derived from known rate constants.

StochSim works by quantising time into a series of discrete, independent time intervals, the sizes of which are determined by the most rapid reaction in the system. In each time interval, a molecule and another object (either a molecule or a pseudo-molecule) is selected at random. If two molecules are selected, any reaction that occurs is bimolecular, whereas if one molecule and a pseudo-molecule are selected, it is unimolecular. Another random number is then generated to determine if a reaction will occur. The probability of a reaction is retrieved from a look-up table and if this exceeds the random number, the particles do not react. On the other hand, if the probability is less than the random number, the particles react, and the system is updated.

StochSim is likely to be slower than the Gillespie algorithm especially when the number of molecules is large. However, if the system contains molecules that can exist in multiple states, then **StochSim** may not only be faster but also closer to physical reality. **StochSim** has been extended to incorporate explicit spatial representation in which nearest-neighbour interactions of molecules (such as clustered receptors on a membrane) can be simulated (Shimizu et al., 2000).

3.1 Gillespie's exact algorithm

To understand the SSA in more detail, we first introduce the *reaction probability density function* $P(\tau, \mu|\mathbf{X})$ defined such that $P(\tau, \mu|\mathbf{X})\mathrm{d}\tau \equiv$ probability that given the state $\mathbf{X}$ at time t, the *next* reaction in the volume will occur in the infinitesimal time interval $(t+\tau, t+\tau+\mathrm{d}\tau)$ *and* will be an R_μ reaction.

The algorithm works by commencing at t_0 with some initial state and randomly picking the time and type of the next reaction to occur from the distribution

$P(\tau, \mu|\mathbf{X}(t_0))$. It then updates the overall state of the system to take account of an occurrence of R_μ and repeats the whole procedure this time picking from the distribution derived from the newly updated state, i.e. $P(\tau, \mu|\mathbf{X}(t_1))$. This process loops repeatedly and in doing so effectively steps through time forming a complete evolution of the system based on the probabilistic model.

To find an expression for $P(\tau, \mu|\mathbf{X})$ we note that it is equal to the probability of *no reaction* over time interval $(t, t+\tau)$, $P_0(\tau|\mathbf{X})$ multiplied by the probability that R_μ will occur over time interval $(t+\tau, t+\tau+\mathrm{d}\tau)$, namely $a_\mu \mathrm{d}\tau$. Thus

$$\begin{aligned} P(\tau, \mu|\mathbf{X}) &= P_0(\tau|\mathbf{X}) a_\mu \mathrm{d}\tau \\ &= P_0(\tau|\mathbf{X}) h_\mu c_\mu \mathrm{d}\tau \ . \end{aligned}$$

It turns out that $P_0(\tau|\mathbf{X})$ has the form (Gillespie, 1977)

$$P_0(\tau|\mathbf{X}) = \exp(-\sum_{\nu=1}^{M} a_\nu \tau) \tag{5}$$

from which we may conclude that

$$P(\tau, \mu|\mathbf{X}) = \begin{cases} a_\mu \exp(-a_0 \tau) \text{ if } 0 \leq \tau < \infty \text{ and } \mu = 1, \cdots, M \\ 0 \qquad\qquad\qquad \text{otherwise} \ , \end{cases}$$

where $a_\mu = h_\mu c_\mu$ and $a_0 = \sum_{\nu=1}^{M} a_\nu$. By noting that $P(\tau, \mu|\mathbf{X})$ is separable, i.e. the product of two functions $f(\mu)$ and $g(\tau)$ which each only depend on one of our two parameters, we see that at any point we can pick τ and μ from the distribution $P(\tau, \mu|\mathbf{X})$ by choosing two random numbers r_1 and r_2 from the interval $[0, 1]$ and setting μ and τ such that

$$\tau = \frac{1}{a_0} \ln \frac{1}{r_1} \tag{6}$$

$$\sum_{\nu=1}^{\mu-1} < r_2 a_0 \leq \sum_{\nu=1}^{\mu} \ . \tag{7}$$

So in summary, after setting the initial species populations $\mathbf{X}$ and reaction constants c_μ the algorithm loops the following steps:

(1) Calculate $a_\mu = h_\mu c_\mu$ $(1 \leq \mu \leq M)$.
(2) Generate r_1 and r_2 and calculate τ and μ according to (6) and (7).
(3) Increase t by τ and adjust $\mathbf{X}$ to take account of an occurrence of R_μ.

3.2 Computational implementation

By following the simple procedure outlined above, a computational algorithm was set up to explore the behaviour predicted by the stochastic approach and compare it to the predictions of the deterministic rate laws. The algorithm implemented the Michaelis–Menten mechanism (1) whereby enzyme and substrate molecules combine to form a complex which can then either disassociate back into the original enzyme-substrate pair or instead convert the substrate molecule into a product molecule leaving the enzyme free to form a new pairing.

In the deterministic approach, the three reactions are controlled by the reactant concentrations X_i and the rate constants k_1, k_{-1} and k_2. By contrast, in the stochastic approach at any given time the reaction *probabilities* are governed by the reactant concentrations X_i and rate constants c_1, c_{-1} and c_2 as described in Section 2.2. It turns out that c_μ and k_μ are essentially equivalent differing only by factors of V, the volume of the vessel (Gillespie, 1977). [2]

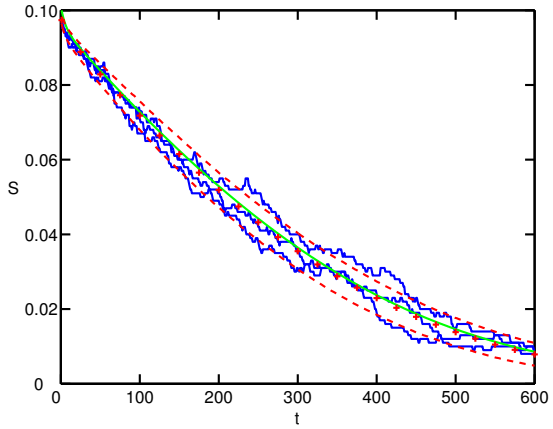
3.2.1 Results and analysis

Thus it is relatively straightforward to contrast the results of the two methods. FIGURE 1(a) shows the results of 2 000 runs of the stochastic algorithm simulating a system with initial molecular populations $S_0 = 100$, $E_0 = 10$, $C_0 = P_0 = 0$ and a volume of 1 000 units [3] . The blue/broken solid curves of each run (3 runs are shown) of the stochastic method show significant random fluctuations from the mean (shown in red/plus sign symbols).

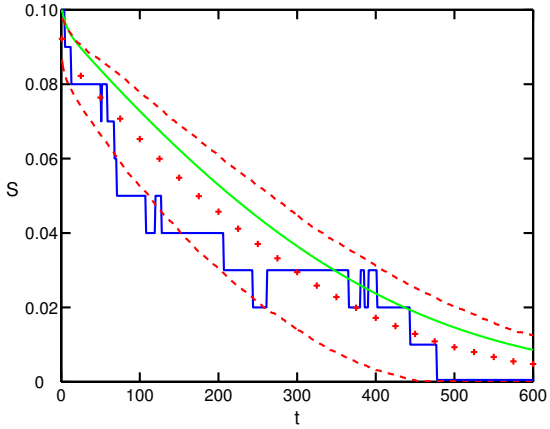
The result of numerically integrating the equations of the deterministic approach is shown in green/smooth solid curve. It is clear that there is a close correspondence between the predictions of the deterministic approach and the stochastic approach, with the deterministic curve falling well within 1 standard deviation (SD) of the stochastic mean (the red/dotted curve). This is a very close match, especially considering our stochastic simulation is modelling a system containing just 110 molecules — well within what we might consider to be the *microscopic* domain.

[2] This is simply a consequence of the way the different constants are defined, with c_μ based on absolute molecular *populations* and k_μ based on molecular *concentrations*.

[3] The unit of volume is arbitrary as long as we are consistent in our units when considering molecular concentrations.



(a) Results for initial molecular populations $S_0 = 100$ and $E_0 = 10$. Three individual simulations are shown explicitly.



(b) Results for initial molecular populations $S_0 = 10$ and $E_0 = 1$. One individual simulation is shown explicitly.

Fig. 1. Stochastic algorithm simulation for the substrate density S in the Michaelis–Menten reaction (1). The blue/broken solid curves are individual simulations. The green/smooth solid curves are the results of numerically integrating the deterministic differential equation. The red/plus sign symbols are the mean for the 2 000 runs of the stochastic simulation and the red/dotted line corresponds to the mean plus (or minus) one standard deviation. Initial conditions are: $C = P = 0$.

However it is worth bearing in mind that an actual in vivo biochemical reaction would follow just one of the many random curves (shown in blue/broken solid curves) that average together producing the closely fitting mean. This curve may deviate significantly from that of the deterministic approach, and thus call into question its validity. Hence, it is perhaps most important to consider the *variance* of the stochastic approach — with a larger variance indicating a greater deviation from the mean and hence from the deterministic curve.

FIGURE 1(b) shows the results for exactly the same simulation setup, except this time we are modelling a system consisting of just 11 molecules within a volume of 100 units [thus the molecular concentrations are equal to those in FIGURE 1(a)]. We see that the deterministic curve (green/smoth solid curve) now shows significant deviation from the mean curve (red/plus sign symbols) but still lies within the 1 SD envelope. However this envelope is now very much wider, indicating that the results of individual runs (blue/broken solid curve) will differ more significantly from the deterministic solution.

Finally, to confirm compatibility of the two approaches FIGURE 2 illustrates how, on average, the stochastic approach tends to the same solution as the deterministic approach as the number of molecules in the system increases, and we hence move from

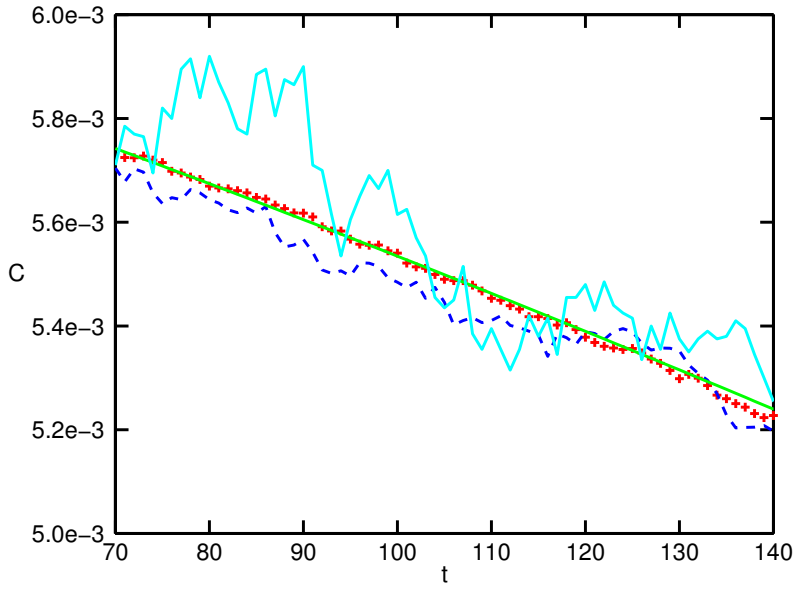


Fig. 2. Mean results from 2 000 runs of the stochastic algorithm simulating systems with varying molecular populations for the enzyme substrate complex C population in the Michaelis–Menten reaction (1). The green/smooth solid line is the numerical solution for the deterministic differential equations. The total number of molecules in the simulation increases from 10 molecules (cyan/broken solid curves), 110 molecules (blue/dash curve) to 1100 molecules (red/plus sign symbols). The initial molecular *concentrations* for the simulations are: $S_0 = 0.10$, $E_0 = 0.01$, $C_0 = P_0 = 0$ molecules per unit volume.

the *microscopic* to the *macroscopic* domain. Coupled with this, we also find (Figure 3) that the $\log SD$ of the data from the 2 000 simulation drops highly linearly as the simulation volume is increased (keeping molecular *concentrations* constant), meaning that each specific run is individually in closer and closer agreement with the deterministic approach as the number of molecules in the system increases. This is a direct effect of the inherent averaging of macroscopic properties of a system of many particles.

These results provide clear verification of the compatibility of the deterministic and stochastic approaches, but importantly also illustrate the validity of the deterministic approach in systems containing as few as 100 molecules. As is clear from Figure 1(a) the match between individual runs of the stochastic simulation and the deterministic solution is still good even for such a small system.

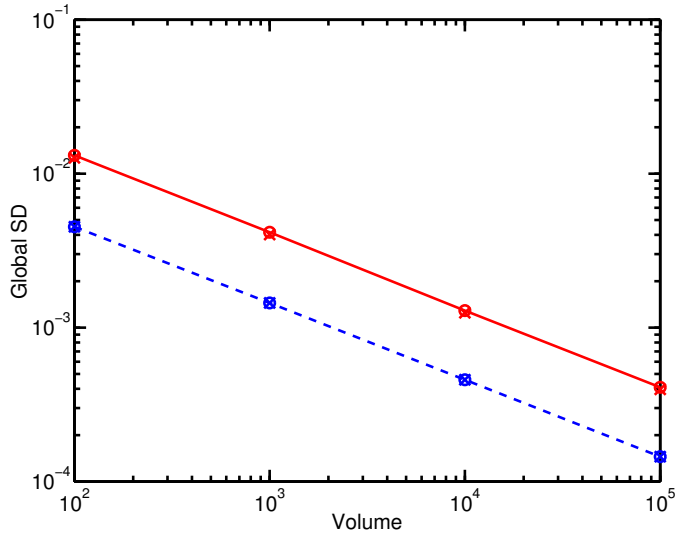


Fig. 3. "Global SD's" (i.e. the mean over the whole simulation time of the S.D. at each time-point) of molecular concentration data from 2 000 runs of the stochastic algorithm for four different simulation volumes with equal initial molecular concentrations. Data for S (circle symbol) and P (cross symbol) are indistinguishable (blue/dash curve), as are data (red/solid curves) for C (circle symbol) and E (cross symbol).

3.3 Enhanced stochastic simulation techniques

Gillespie's algorithm suffers from a rapidly increasing computational overhead as the complexity of the system being modelled is increased. It is very common in biochemistry to have systems with several or even tens of chemical reactants interacting via an array of distinct reaction mechanisms. The key point about the SSA is that the time step τ must be small enough to guarantee that only one reaction occurs in that time interval, and as such, increasing the molecular population or number of reaction mechanisms necessarily requires a corresponding decrease in τ. Clearly the SSA can be very computationally inefficient especially when there are large numbers of molecules or the propensity functions are large.

Now if the system possesses a macroscopically infinitesimal timescale so that during any dt all of the reaction channels can fire many times, yet none of the propensity functions change appreciably, then the discrete Markov process as described by the SSA can be approximated by a continuous Markov process. This Markov process is described by the Chemical Langevin Equation (CLE), which is a stochastic ordinary differential equation (SDE) – see Gillespie (1992b). Thus the vector of chemical species as a function of time can be viewed as the solution of an SDE in Itô form

$$\mathrm{d}X = \sum_{j=1}^{M} \nu_j a_j(X)\mathrm{d}t + \sum_{j=1}^{M} \nu_j \sqrt{a_j(X)}\mathrm{d}W_j(t) \ , \tag{8}$$

where the $W_j(t)$ are independent Wiener processes.

The CLE represents processes in the intermediate regime, that is those processes that are stochastic and continuous. A Wiener process is a stochastic process satisfying

$$E(W(t)) = 0, \quad E(W(t)W(s)) = \min\{t, s\} \ .$$

It is known that the Wiener increments are independent Gaussian processes with mean 0 and variance $|t - s|$ (that is, $N(0, |t - s|)$). Thus the Wiener increment $\Delta W(t) \equiv W(t+\Delta t) - W(t)$ is a Gaussian random variable $N(0, \Delta t) = \sqrt{\Delta t} N(0, 1)$.

The Chemical Langevin Equation is an example of the more general class of Itô Stochastic Differential Equations given by

$$dy(t) = g_0(y(t))dt + \sum_{j=1}^{d} g_j(y(t))\, dW_j(t), \qquad y(t_0) = y_0, \quad y \in \mathbb{R}^m. \tag{9}$$

Thus, general classes of methods that can be used to solve (9) can also be used to simulate solutions of (8) (see, for example, Kloeden and Platen, 1992).

We note that the third regime occurs when the noise terms are negligible compared with the deterministic term. This leads to the standard chemical kinetic approach that is described by the reaction rate equations

$$X'(t) = \sum_{j=1}^{M} \nu_j a_j(X(t)) \ .$$

There are standard ODE techniques for solving such a system. The efficacy of such methods depends on whether the system is stiff or not — that is whether or not there are widely differing time constants. If there are, then explicit methods cannot be used and A-stable implicit methods are required.

Recently, considerable attention has been paid to reducing the computational time of simulation algorithms for stochastic chemical kinetics. Gibson and Bruck (2000) refined the first reaction SSA of Gillespie by reducing the number of random variables that need to be simulated. This can be effective for systems in which some reactions occur much more frequently than others. Resat et al. (2001) treat systems which have widely varying rate constants by applying a weighted Monte Carlo approach.

Gillespie (2001) proposed two new methods, namely the τ-leap method and the

midpoint τ-leap method in order to improve the efficiency of the SSA while maintaining acceptable losses in accuracy. The key idea here is to take a larger time step and allow for more reactions to take place in that step, but under the proviso that the propensity functions do not change too much in that interval. Thus in the time interval $[t, t+\tau)$ and with the present state $X(t)$ at time t, then the number of times that the reaction channel R_j will fire is a Poisson random variable

$$K_j(\tau; X, t) = P(a_j(X), \tau), \quad j = 1, \ldots, M.$$

Here the notation $P(\lambda, t)$ denotes a stochastic Poisson process with mean λt and variance λt and where

$$\Pr(P(\lambda, t) = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!} \ .$$

These considerations lead to the τ-leap method.

3.3.1 The τ-leap method

Choose a value for τ that satisfies the Leap Condition: i.e., a temporal leap by τ will result in a state change λ such that for every reaction channel R_j, $|a_j(X+\lambda) - a_j(X)|$ is "effectively infinitesimal". Generate for each $j = 1, \ldots, M$ a sample value k_j of the Poisson random variable $P(a_j(X), \tau)$, and compute $\lambda = \sum_{j=1}^{M} k_j \nu_j$. Finally, perform the updates by replacing t by $t + \tau$ and X by $X + \lambda$.

Burrage and Tian (2003) introduced the framework of Poisson Runge-Kutta (PRK) methods for simulating chemical reaction systems. These PRK methods are related to the class of stochastic Runge-Kutta (SRK) methods for solving stochastic differential equations driven by Wiener noise.

The reason for adopting this framework is as follows. A Poisson random variable $P(a_j(X), \tau)$ with a large mean $a_j(X)\tau$ can be approximated by a Gaussian random variable $N(a_j(X)\tau, a_j(X)\tau)$, since

$$P(a_j(X), \tau) \approx N(a_j(X)\tau, a_j(X)\tau) = a_j(X)\tau + \sqrt{a_j(X)\tau} N(0, 1) \ ,$$

where $N(\mu, \sigma^2)$ is a Gaussian random variable with mean μ and variance σ^2. This can be viewed as

$$P(a_j(X), \tau) \approx a_j(X)\tau + \sqrt{a_j(X)} \Delta W(t) \ . \tag{10}$$

Now the simplest numerical method for solving (9) is the Euler-Maruyama method. It takes the form

$$y_{n+1} = y_n + hg_0(y_n) + \sum_{j=1}^{d} \Delta W_j^{(n)} g_j(y_n), \quad t_{n+1} = t_n + h,$$

where $\Delta W_j^{(n)} \equiv W_j(t_n + h) - W_j(t_n)$ is a Gaussian random variable $N(0, h)$.

The Euler-Maruyama method converges with strong order 0.5 and weak order 1 to the Itô form of the SDE. If it is applied to (8) it takes the form

$$X_{n+1} = X_n + \tau \sum_{j=1}^{M} \nu_j a_j(X_n) + \sum_{j=1}^{M} \Delta W_j^{(n)} \nu_j \sqrt{a_j(X_n)} \ .$$

Now using the approximation in (10) we can write this as

$$X_{n+1} = X_n + \sum_{j=1}^{M} \nu_j P_j(a_j(X_n), \tau) \ . \tag{11}$$

This method is nothing but the τ-leap method of Gillespie. Thus the τ-leap method is the Euler method applied in the discrete setting when there are small numbers of molecules. This means that we can essentially apply the same algorithm in different regimes, which is important in attempting to use multi-scaled, partitioning techniques – see below.

More recently, Tian and Burrage (2004a) have considered sampling from a Binomial distribution rather than a Poisson distribution in (11). This avoids generating negative molecular numbers that can occur with the Poisson leap methods and appears to lead to more robust methods with significant improvements in accuracy and efficiency.

Rathinam et al. (2003) consider how stiffness manifests itself at both the continuous deterministic and discrete stochastic levels. In this case explicit methods become impractical. The authors construct two implicit versions of the explicit τ-leap method known as the rounded and unrounded implicit τ-leap method, which have better stability properties than the explicit τ-leap method and are suitable for solving stiff chemical systems. The rounded method has the form

$$\begin{aligned} X &= X_n + \tau \sum_{j=1}^{M} \nu_j \left(a_j(X) - a_j(X_n)\right) + \sum_{j=1}^{M} \nu_j P_j \left(a_j(X_n), \tau\right) \\ X_{n+1} &= X_n + \sum_{j=1}^{M} \nu_j \left[\tau \left(a_j(X) - a_j(X_n)\right)\right] + \sum_{j=1}^{M} \nu_j P_j \left(a_j(X_n), \tau\right) \ , \end{aligned}$$

where [ ] denotes the nearest nonnegative integer.

4 Incorporating the Quasi-Steady-State Assumption in the stochastic formulation

One of the great challenges to the efficient simulation of chemical kinetic systems is how we deal with mixed systems in which some key species have low abundances (as is the case for some molecules in genetic regulation) while other molecules have large abundances and can be modelled via continuous SDEs. Thus a vital question to address is how we can link discrete and continuous models and simulation algorithms in a sensible and efficient manner when treating chemical kinetic systems?

Recently two new approaches by Haseltine and Rawlings (2002) and Rao and Arkin (2003) have been proposed in an attempt to speed up the performance of the SSA. Both of these ideas are based on partitioning of the system. In the case of Rao and Arkin, they consider a timescale separation in which a subset of the system is asymptotically at steady state. This is called the quasi-steady-state assumption (QSSA) and eliminates the fast dynamics that is responsible for the poor computational performance of the SSA. The QSSA is a simplification derived from the deterministic approach to reduce the number of coupled differential equations governing the dynamics of the system under study (see Schnell and Maini, 2003, for a review). It assumes that one or more intermediate molecules within a chemical system quickly reach a quasi-equilibrium state whereby their rate of formation and destruction approximately sum to zero. Hence, applying the QSSA to deterministic kinetics, results in the ODEs describing the intermediate species being set to 0 (Schnell and Mendoza, 1997; Schnell and Maini, 2000). Similarly, in the stochastic setting the system is split into primary (y) and ephemeral (z) subsystems.

Inherent in the QSSA is the assumed macroscopic nature of the system allowing the averaging out of microscopic fluctuations in molecular populations.

Accordingly, Haseltine and Rawlings (2002) attempt to speed up the performance of the SSA by partitioning a chemical reaction system into slow and fast reaction subsets. The slow subsystem corresponds to extents with small propensity functions and few numbers of reactions, while the latter corresponds to large propensity functions and large numbers of reactions. This partitioning is achieved by exploiting the structure of the CME and deriving master equations that describe the evolution of the probability density function for both the slow and fast subsystems. The slow

system is treated by the SSA, while the fast system is treated either deterministically or by applying the explicit Euler-Maruyama method to the CLE. Thus at each time point t_n the CLE is repeatedly solved until $t_{n+1} = t_n + \tau$ is reached, then the SSA is applied to the slow subsystem with a stepsize of τ.

Burrage et al. (2004) extended these ideas to classify a system (in terms of both the size of the propensity functions and the number of molecules in the system) into slow, intermediate and fast reactions. They form three vectors corresponding to the slow, intermediate and fast regimes and place in those vectors the corresponding reaction numbers. If there are no reactions in, say, the intermediate vector for a given time step, this means there are no intermediate reactions for that step and the simulation regime changes accordingly. They use the SSA, the τ-leap method, and the Euler-Maruyama method in the slow, intermediate and fast regimes, respectively.

Returning to the Rao and Arkin approach, we explicitly consider the important case of the Michaelis–Menten mechanism (1). Subject to the condition derived analytically by Schnell and Mendoza (1997)

$$1 \gg \frac{E_0}{K_M + S_0} \approx \frac{E_0}{S_0} \text{ if } k_1 \gg k_{-1}, k_2 \tag{12}$$

where $K_M = (k_{-1}+k_2)/k_1$, the QSSA assumes after some initial transient period that the complex concentration C stays constant, i.e. $\mathrm{d}C/\mathrm{d}t \approx 0$. This reduces the number of coupled differential equations by one as well as simplifying those that remain. Rao and Arkin (2003) propose that application of the QSSA to the stochastic formulation has a similar simplifying effect by excluding the quasi-steady-state *"intermediate"* chemical species from the state vector $\mathbf{X}$ thus reducing its dimensionality. This in turn reduces the dimensionality and complexity of the chemical master equation of the system (4).

By splitting state vector $\mathbf{X}$ into the primary species vector $\mathbf{Y}$ and the intermediate species vector $\mathbf{Z}$ such that $\mathbf{X} \equiv (\mathbf{Y}, \mathbf{Z})$, the chemical master equation can be rewritten

$$\frac{\partial P(\mathbf{Y}, \mathbf{Z}; t)}{\partial t} = \sum_{\mu=1}^{M} a_\mu(\mathbf{Y} - \mathbf{v}_\mu^Y, \mathbf{Z} - \mathbf{v}_\mu^Z) P(\mathbf{Y} - \mathbf{v}_\mu^Y, \mathbf{Z} - \mathbf{v}_\mu^Z; t) - a_\mu(\mathbf{Y}, \mathbf{Z}) P(\mathbf{Y}, \mathbf{Z}; t) \ . \tag{13}$$

We then make use of the definition of conditional probabilities

$$P(\mathbf{Y}, \mathbf{Z}; t) = P(\mathbf{Z}|\mathbf{Y}; t) P(\mathbf{Y}; t)$$

and the chain rule of differentiation to rewrite (13) as

$$P(\mathbf{Y};t)\frac{\partial P(\mathbf{Z}|\mathbf{Y};t)}{\partial t}+P(\mathbf{Z}|\mathbf{Y};t)\frac{\partial P(\mathbf{Y};t)}{\partial t}=\sum_{\mu=1}^{M}[a_\mu(\mathbf{Y}-\mathbf{v}_\mu^Y,\mathbf{Z}-\mathbf{v}_\mu^Z)$$
$$\times P(\mathbf{Z}-\mathbf{v}_\mu^Z|\mathbf{Y}-\mathbf{v}_\mu^Y;t)P(\mathbf{Y}-\mathbf{v}_\mu^Y;t)$$
$$-a_\mu(\mathbf{Y},\mathbf{Z})P(\mathbf{Y},\mathbf{Z};t)]\ .$$

We then apply the QSSA by setting the net rate of change of the conditional probability of the intermediate species to zero

$$\frac{\partial P(\mathbf{Z}|\mathbf{Y};t)}{\partial t}\approx 0 \tag{14}$$

eliminating the first term from the LHS of the above equation. By noting $\sum_z P(\mathbf{Z}|\mathbf{Y};t)=1$ we arrive at the *approximate master equation*, which is dependent only on $\mathbf{Y}$

$$\frac{\partial P(\mathbf{Y};t)}{\partial t}=\sum_{\mu=1}^{M}[b_\mu(\mathbf{Y}-\mathbf{v}_\mu^Y)P(\mathbf{Y}-\mathbf{v}_\mu^Y;t)-b_\mu(\mathbf{Y})P(\mathbf{Y};t)] \tag{15}$$

where

$$b_\mu(\mathbf{Y})=\sum_z a_\mu(\mathbf{Y},\mathbf{Z})P(\mathbf{Z}|\mathbf{Y}).$$

Thus we have an approximate stochastic form for the dynamics of the primary species dependent only on the state of the primary species. From this point we are able to use a modified version of Gillespie's algorithm to go from the chemical master equation to the generation of a stochastic simulation of the system. The modification involves at each time step picking $\mathbf{Z}$ from the conditional probability function $P(\mathbf{Z}|\mathbf{Y})$, before using it to calculate $a_\mu(\mathbf{Y},\mathbf{Z})\equiv a_\mu(\mathbf{X})$ for the M reactions in the usual way.

Taking as an illustrative example the Michaelis–Menten mechanism (1), Rao and Arkin (2003) show that by application of the QSSA choosing

- Primary species: $S_T=S+C\equiv$ total concentration of substrate, free and bound
- Intermediate species: $C\equiv$ enzyme-substrate complex concentration

we can consider the simplified system

$$S\rightarrow P \tag{16}$$

with associated chemical master equation

$$\frac{dP(S_T;t)}{dt}=a(S_T+1)P(S_T+1;t)-a(S_T)P(S_T;t)\ . \tag{17}$$

In this equation,

$$a(S_T) = \frac{v_{\max} S_T}{K_M + S_T} \ ,$$

where $v_{\max} = k_2 E_0$ is the maximum velocity of the reaction and $K_M = (k_{-1} + k_2)/k_1$ is the Michaelis–Menten constant. Thus we have a much simplified chemical master equation involving just one (as opposed to the original three) reaction mechanism, and in principle stochastic simulation of this system becomes less computationally demanding. This technique of combining more than one species (S and C) into a single aggregate variable (S_T) is known as *lumping*, and as shown by (Schnell and Maini, 2002) is the basis of the total Quasi-Steady-State Assumption (tQSSA) – an alternative approximation technique valid in many cases where the standard QSSA is not.

However, the clear computational benefit of this simplification is offset to a degree by the need to calculate or approximate the conditional probability function $P(\mathbf{Z}|\mathbf{Y}) \equiv P(C|S_T)$. In this case, rather than generating and randomly selecting from the probability function $P(C|S_T)$ at each time step, an analytical expression for the expectation $E(C|S_T)$ is instead used.

Rao and Arkin further apply the QSSA to more complex biochemical systems involved in gene regulation, illustrating the computational savings inherent in their approach. In doing so they reduce the number of reactions considered in the system from 10 down to 2. In this case, rather than explicitly calculating conditional probability function $P(\mathbf{Z}|\mathbf{Y})$ it is instead approximated with a Gaussian distribution. The modified Gillespie algorithm is then implemented and shown to cause a 50% reduction in computational time with minimal loss of accuracy.

Whilst this evidence is convincing as to the accuracy and computational benefit of Rao and Arkin's technique, it is important to consider how valid it is to apply the QSSA within the stochastic method – a method whose benefit over the deterministic approach has been shown to exist only within the microscopic domain. The QSSA is fundamentally derived from the deterministic approach where large numbers of molecules are assumed. As such, it is far from obvious that we can safely transfer this approximation to the microscopic domain. Thus it seems pertinent to ask whether it is valid at all to incorporate the QSSA within the stochastic framework.

5 Two-dimensional Monte Carlo simulations

Common to all the stochastic methods considered so far is the assumption that the chemical system is *well mixed* at all times. This allows us to make the simplifying assumption that throughout the reaction any given particle has equal chance of colliding with any other particle wherever in the volume they are each located.

This *well mixed* assumption requires that the diffusion of molecules through the volume be sufficiently unrestricted that the average time taken for a molecule to go from one reactive collision to another is significantly greater than the average time taken to diffuse across the volume.

The structural organisation of the cytoplasm has only recently come to the fore in modelling in vivo kinetics (Ellis, 2001). Intracellular environments are characterised by significant physical structures that are likely to seriously inhibit the diffusion of molecules across the reaction volume. As such the validity of the stochastic approaches discussed so far is called into question when considering in vivo reactions (see Schnell and Turner, 2004, and references therein).

Chemical reactions in crowded environments show fractal-like kinetic properties (Kopelman, 1986, 1988). Berry (2002) and Schnell and Turner (2004) implemented a lattice gas automata method using a Monte Carlo algorithm on a two dimensional square lattice with cyclic boundary conditions. Each molecule is mobile on the lattice through diffusion, modelled by independent nearest-neighbour random walks of the individual molecules. Time is split into discrete steps and at each step molecules are selected at random to take a single step in a random direction along the grid. In this way, molecules move through the volume via two-dimensional random walks known as blind ant processes (Majid et al., 1984). When two compatible molecules collide, they react with a certain probability. For enzyme-catalysed reactions, enzyme-substrate complex molecules when randomly chosen may also spontaneously disassociate with a set probability. Obstacles are represented as an extra, non-reactive, immobile molecular type. The coordinates of the position of every molecule and occupancy status of each lattice site are stored and used for analysis. At any moment of the simulation, one given lattice site cannot be occupied by more than one molecule.

In our simple model of the Michaelis–Menten reaction (1), the rate coefficients k_1, k_{-1} and k_2 are modelled by the reaction probabilities f, r and g respectively where

- f is the probability that an enzyme and substrate molecule will react to form a complex molecule ($E + S \rightarrow C$) given that they have collided on the lattice.
- r and g are the probabilities that when randomly selected by the Monte Carlo method, a given complex molecule will disassociate into respectively an enzyme and substrate molecule ($C \rightarrow E + S$) and an enzyme and product molecule ($C \rightarrow E + P$).

Thus, we expect for a sufficiently large system, or for the averaging of the data from sufficient repetitions of the simulation, that the numerically derived rate coefficients k_1, k_{-1} and k_2 will respectively tend to the reaction probabilities f, r and g.

This simulation method differs from the stochastic simulation algorithm of Gillespie (1977) in one crucial way: Gillespie's Exact Algorithm utilises the spatial homogeneity of the reaction environment to derive a probability distribution for the time between elementary reaction events. The algorithm then samples randomly from this distribution to simulate the dynamics of the reaction. In contrast, the Monte Carlo method assumes only that the molecular motion is Brownian in nature – i.e. that it progresses via individual molecular random walks over a discrete lattice. As such, the time between second order reaction events does not conform to a predetermined probability distribution, but rather is dictated by two factors: firstly, the chance occurrence of two random walks bringing together two molecules on the same site simultaneously. Secondly, the preset reaction probabilities dictating how likely two co-incident molecules are to react.

At the beginning of each simulation, the E and S molecules and the obstacles (if present) are placed on the lattice by randomly choosing the co-ordinates for each of them. At each Monte Carlo sequence, a "subject" molecule is chosen at random and moved/reacted upon according to the following rules:

1. Randomly choose nearest neighbour "destination" site.
2. If the "subject" molecule is E, S or P and destination site is empty, move to it.
3. Otherwise

3.1. If the "subject" molecule is E or S and the molecule occupying the "destination" site ("target" molecule) is respectively S or E, then generate a random number between 0 and 1. If this is lower than reaction probability f, replace the "target" molecule with C, remove the "subject" molecule and set $\gamma = \gamma + 1$, where $\gamma(t)$ is simply a counter of the number of $E + S \rightarrow C$ reactions that have occurred in time interval $[0, t]$.

3.2. If the "subject" molecule is C, check vacancy of nearest neighbour sites. If at least one nearest neighbour site is vacant, randomly choose a vacant "destina-

tion" nearest neighbour site and generate a random number x between 0 and 1.

3.2.1. If $x < r$ place E on the "subject" site and S on the "destination" site. This step corresponds to the elementary reaction $C \rightarrow E + S$.

3.2.2. If $r \leq x < r + g$ place E on the "subject" site and P on the "destination" site. This step corresponds to the elementary reaction $C \rightarrow E + P$.

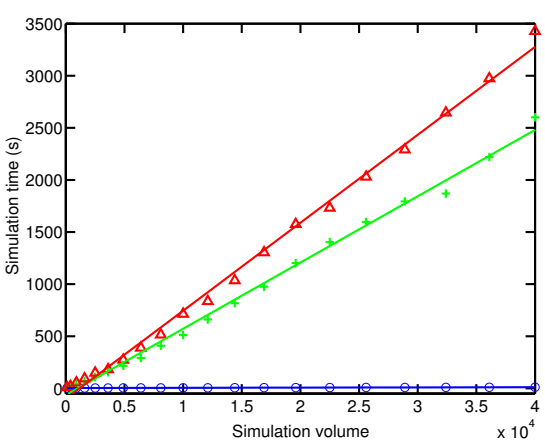
3.2.3. If $x \geq r + g$ move C to the "destination" site.

4. Otherwise, keep "subject" molecule on initial site. No movement or reaction occurs.

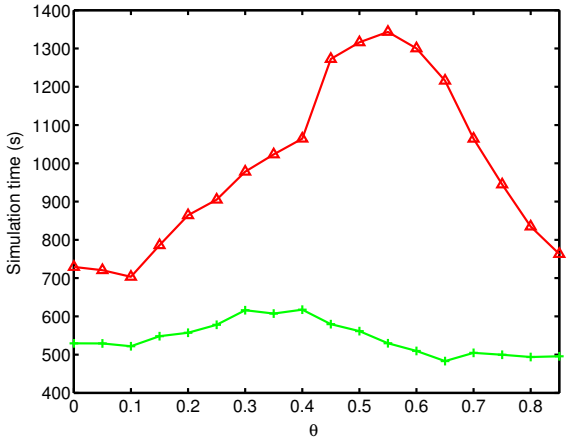
For each time step, the Monte Carlo sequence is repeated $n_{\text{total}}(t)$ times, where $n_{\text{total}}(t)$ is the number of distinct molecules on the lattice (excluding obstacles) at time t. Despite a fully conservative system where no molecules are created or destroyed, the total number of distinct molecules n_{total} changes over time because when an E and an S molecule combine to form a C molecule, we have one fewer distinct molecule in the system. Setting this number of repetitions ensures that one time unit statistically represents the time necessary for each molecule to move once. The simulation proceeds until a predetermined final time point.

The net rates for bimolecular reactions, averaged over the spatial grid, have been found to decrease with time, following an empirical time-dependent relationship $k(t) = k_0(\tau + t)^{-h}$, where k_0 is the ideal (dilute solution) rate constant, and the positive parameters h and τ are found to depend on the number and arrangement of the obstacles (Schnell and Turner, 2004). In diffusion-limited reactions the rate depends on the geometry of the obstacles, leading to fractal-like effects in the reaction rates that are hardly separable from the purely geometric effects (Kopelman, 1986). Numerous studies have been made to tabulate the values of fractal-like scaling exponents for reactions in different geometries (see, for example Ahn et al., 1999, and reference therein).

The lattice gas approach offers the significant benefit of being able to incorporate a non-homogenous environment. However, this benefit comes at a large computational cost. FIGURE 5(a) shows the time taken to perform 10 runs of the SSA (blue/circle symbols), a two-dimensional lattice gas algorithm (green/plus sign symbols) and a three-dimensional lattice gas algorithm (red/triangle symbols) for various sizes of simulation environment on a 2.8MHz Intel Pentium4 running **MATLAB** within *Microsoft* **Windows XP**. For the lattice gas simulations, the term "volume" refers to the total number of grid elements present on the two-dimensional or three-dimensional lattice.



(a) Average simulation time in seconds for various sizes of the simulation environment.



(b) Average simulation time in seconds as a function of obstacle density, θ, for the lattice gas simulations.

Fig. 4. Benchmarking results for the stochastic simulations. We compared the simulation time (seconds) for the Gillespie algorithm simulations (blue/circle symbols), two-dimensional (green/plus sign symbols) and a three-dimensional (red/triangle symbols) lattice gas algorithms.

As is evident from the figure, all three approaches scale approximately linearly with volume, but with the lattice gas approaches growing in cost at at a far higher rate than the SSA. The three-dimensional lattice gas simulations have an additional premium due to the increased overhead of database searching through three-dimensional rather than two-dimensional data arrays. It is also worth noting when comparing two- and three-dimensional systems that for a given volume they have the same number of elements, so the three-dimensional cube will be of much smaller side length than its equivalent two-dimensional square.

FIGURE 5(b) shows the variation in computational load for the two- (green/plus sign symbols) and three-dimensional (red/triangle symbols) lattice gas approaches as a function of obstacle density, θ. Evidently, the computational overhead increases with θ as expected until approximately $\theta = \theta_c$ (~ 0.5) at which point the load tails off. This drop off is likely to be due to the formation of isolated sub-volumes within the reaction environment when $\theta > \theta_c$, thus reducing the frequency of reaction events and consequently the computational load.

We must also acknowledge that the lattice gas simulations represent a simplified picture of the cellular environment. They model reactions in fully conservative conditions where no external factors have an effect on the observed kinetics. They also allocate structure to the environment in an entirely random fashion, in contrast to the highly organised structure of living cells (Kuthan, 2001). The lattice gas au-

tomata algorithm has restricted the motion of our reactant molecules to discrete jumps in restricted directions where in reality, the movement of these reactants will be entirely continuous. These restrictions may potentially have unforseen effects on the overall dynamics of the simulated reaction in terms of the size of the lattice and its mesh size. Furthermore, from a physical point of view, the simulations conserved momentum within the cell only on average, with individual reactants changing direction entirely independently. Thus it is important to verify the extent to which the results of the lattice gas automata conform to well understood in vivo experimental results.

6 Discussion and conclusions

Several levels of detail have traditionally been employed in modelling biochemical reactions and pathways (Crampin et al., 2004). The most detailed level of description is the chemical kinetics approach, in which the concentrations (or numbers) of the molecules involved in reactions are modelled over time. Normally, the kinetic models consist of a system of ODEs that can be analysed with nonlinear dynamics techniques and numerically computed with standard software packages. Unfortunately, the representation of a biochemical reaction as system of ODEs is completely deterministic and does not take into account the random noise of fluctuations in concentration within the cell.

The stochastic kinetic modelling approach provides a more detailed description for reactions than the systems of ODEs. There are numerous stochastic approaches for modelling reactions, but they are difficult to implement analytically and researchers are reduced to numerical studies (see, for a review Burrage et al., 2004). Another caveat to the use of stochastic modelling is that the appropriate approach to describe the fluctuations of biochemical reactions must be used. There are clearly two sources of fluctuations in in vivo biochemical reactions: (i) small number of molecules – simulated typically in reactions involving protein-DNA binding, transcription and translation – and (ii) limited diffusion effects due to the structural organisation of the cytoplasm and the macromolecular crowding.

The Gillespie approach is an efficient method to model chemical reactions taking into account the effects that only small quantities of molecules are involved. Having introduced the key properties of the stochastic approach to biochemical kinetics and implemented Gillespie's SSA to model the Michaelis–Menten mechanism, we have shown that results from the stochastic and deterministic methods are consistent.

Even in highly microscopic environments, mean data from the SSA (Gillespie, 1976, 1977) was seen to coincide closely with the predictions of the law of mass action confirming that the two approaches remain consistent on average for large and small systems (Gillespie and Mangel, 1981). We have demonstrated this in the case of the Michaelis–Menten mechanism, where the deterministic approach remains in close agreement with the stochastic approach even when considering highly microscopic environments with as few as 100 molecules present. This suggests that when considering reaction kinetics in vivo the deterministic approach can indeed be a valid one to use.

We have discussed the benefits and potential problems of incorporating the QSSA into the stochastic approach, along with attempts to take account of the finite timescale of individual reaction cycles. We have also discussed attempts to improve on algorithmic efficiencies through the τ-leap and more general approaches, sampling either from a Poisson or Binomial distribution. There is strong evidence that sampling from the Binomial distribution confers greater robustness and improved efficiencies.

However such approaches have their own difficulties involving assumptions that molecules act as dimensionless point-particles and environments are entirely homogeneous and well stirred, ensuring the probability of a molecule existing in any sub-region of the vessel is equal across the whole volume. These assumptions render these approaches unsuitable for incorporating the fluctuation effects of limited diffusion due to the cytoplasm structure and macromolecular crowding.

The diffusion-limited reaction kinetics differ significantly from the classical kinetic approaches. In most of the limited diffusion stochastic approaches presented in the literature (see Calef and Deutch, 1983, for a review), the kinetics laws resemble mass action kinetics laws: reactions are driven by time independent rate constants, which are proportional to the microscopic diffusion constants of the reactants. The stochastic approach is based on the mean square displacement, which is linear in time for homogeneous systems. Alternative approaches can be based on the first passage time or on the exploration space (Kopelman, 1986); "fractal-like kinetics" is one such approach. Tian and Burrage (2004b) have also found that the standard rate constant can be inappropriate in certain genetic regulatory networks.

Recently two-dimensional Monte Carlo simulations incorporating immobile obstacles have been introduced to model limited diffusion biochemical reactions (Berry, 2002; Schnell and Turner, 2004). Results from this approach have shown that the kinetics of the reaction are indeed altered by the presence of heterogeneities, and this effect is

well modelled by the "fractal-like kinetics". Thus, Monte Carlo simulation appears to be the most promising stochastic technique for modelling in vivo biochemical kinetics as it can take into account both the small number of reacting molecules and the limited diffusion. This approach can be improved by simulating an environment more closely representative of a living cell and no longer considering the system in isolation. It will certainly yield a detailed picture of the behaviour of a biochemical pathway. However, this completeness would come at a high computational cost and does not provide an analytical treatment to further our understanding of the system under consideration. It is worth making the effort of implementing the SSA to the analytical description of the fractal-like kinetics approach in order to obtain a more computationally efficient approach for modelling reactions in in vivo conditions.

Acknowledgements

TET was the recipient of the EPRSC scholarship during the development of this work. SS has been funded by the Research Training Fellowship Programme in Mathematical Biology (Grant No. 069155) of the Wellcome Trust (London). KB would like to thank the Australian Research Council for funding under the Federation Fellowship scheme. Part of this work has been carried out during a visit by SS to the Bioengineering Institute, University of Auckland (Auckland, New Zealand) in July 2003 and the Imaging Center of the Stowers Institute for Medical Research (Kansas City, Missouri, USA) in September 2003: their hospitality is kindly acknowledged. This document has been partially typed with the aid of peditPro provided by the courtesy of PaulComputing (http://www.paulcomputing.com).

References

Ahn, J., Kopelman, R., Argyrakis, P., 1999. Hierarchies of nonclassical reaction kinetics due to anisotropic confinements. J. Chem. Phys. 110, 2116–2121.

Arányi, P., Tóhn, J., 1977. A full stochastic description of the Michaelis–Menten reaction for small systems. Acta Biochim. et Biophys. Acad. Hung. 12, 375–388.

Arkin, A., Ross, J., McAdams, H. H., 1998. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *Escherichia coli* cells. Genetics 149, 1633–1648.

Bartholomay, A., 1957. A stochastic approach to chemical reaction kinetics. Ph.D. thesis, Harvard University.

Bartholomay, A., 1962a. Enzyme reaction rate theory: a stochastic approach. Ann. N. Y. Acad. Sci 96, 897–912.

Bartholomay, A., 1962b. A stochastic approach to stastical kinetics with application to enzyme kinetics. Biochemistry 1, 223–230.

Berry, H., 2002. Monte Carlo simulations of enzyme reactions in two dimensions: Fractal kinetics and spatial segregation. Biophys. J. 83, 1891–1901.

Blake, W. J., Kærn, M., Cantor, C. R., Collins, J. J., 2003. Noise in eukaryotic gene expression. Nature 422, 633–637.

Burrage, K., Tian, T., 2003. Poisson runge-kutta methods for chemical reaction systems. Proc. Hong Kong Conf. Sci. Comput., in press.

Burrage, K., Tian, T., Burrage, P., 2004. A multi-scaled approach for simulating chemical reaction systems. Prog. Biophys. Mol. Biol., in press.

Calef, D. F., Deutch, J. M., 1983. Diffusion-controlled reactions. Ann. Rev. Phys. Chem. 34, 493–524.

Clegg, J. S., 1984. Properties and metabolism of the aqueous cytoplasm and its boundaries. Am. J. Physiol. 246, R133–R151.

Cox, B. G., 1994. Modern Liquid Phase Kinetics. Oxford University Press, Oxford.

Crampin, E. J., Schnell, S., 2004. New approaches to modelling and analysis of biochemical reactions, pathways and networks. Prog. Biophys. Mol. Biol., in press.

Crampin, E. J., Schnell, S., McSharry, P. E., 2004. Mathematical and computational techniques to deduce complex biochemical reaction mechanisms. Prog. Biophys. Mol. Biol., in press.

Darvey, I. G., Staff, P. J., 1967. The application of the theory of Markov processes to the reversible one substrate–one intermediate–one product enzymic mechanism. J. theor. Biol. 14, 157–172.

Ellis, R. J., 2001. Macromolecular crowding: obvious but underappreciated. Trends Biochem. Sci. 26, 597–604.

Elowitz, M. B., Leibler, S., 2000. A synthetic oscillatory network of transcriptional

regulators. Nature 403, 335–338.

Elowitz, M. B., Levine, A. J., Siggia, E. D., Swain, P. S., 2002. Quantitative analysis of the time courses of enzyme-catalyzed reactions. Science 297, 183–186.

Epstein, I. R., Pojman, J. A., 1998. An introduction to nonlinear chemical dynamics: oscillations, waves, patterns, and chaos. Oxford University Press, Oxford.

Espenson, J. H., 1995. Chemical kinetics and reaction mechanisms. McGraw-Hill, Singapore.

Fedoroff, N., Fontana, W., 2002. Genetic networks: Small numbers of big molecules. Science 297, 1129–1131.

Firth, C. J., Bray, D., 2000. Stochastic simulation of cell signalling pathways. In: Bower, J. M., Bolouri, H. (Eds.), Computational modeling of genetic and biochemical networks. MIT Press, Cambridge, MA, pp. 263–286.

Gibson, M. A., Bruck, J., 2000. Efficient exact stochastic simulation of chemical systems with many species and many channels. J. Phys. Chem. A. 104, 1876–1889.

Gillespie, D. T., 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. Comput. Phys. 22, 403–434.

Gillespie, D. T., 1977. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81, 2340–2361.

Gillespie, D. T., 1992a. Markov processes. Academic Press Inc., Boston, MA.

Gillespie, D. T., 1992b. A rigorous derivation of the chemical master equation. Physica A 188, 404–425.

Gillespie, D. T., 2001. Approximate accelerated stochastic simulation of chemically reacting systems. J. Chem. Phys. 115, 1716–1733.

Gillespie, D. T., Mangel, M., 1981. Conditioned averages in chemical-kinetics. J. Chem. Phys. 75, 704–709.

Gonze, D., Halloy, J., Goldbeter, A., 2002. Robustness of circadian rhythms with respect to molecular noise. Proc. Natl. Acad. Sci. USA 99, 673–678.

Halling, P. J., 1989. Do the laws of chemistry apply to living cells? TIPS 14, 317–318.

Haseltine, E. L., Rawlings, J. B., 2002. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. J. Chem. Phys. 117, 6959–6969.

Heinrich, R., Schuster, S., 1996. The regulation of cellular systems. Chapman and Hall, New York.

Hume, D. A., 2000. Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression. Blood 96, 2323–2328.

Jachimowski, C. J., McQuarrie, D. A., Russell, M. E., 1964. A stochastic approach to enzyme-substrate reactions. Biochemistry 3, 1732–1736.

Kerker, M., 1974. Brownian movements and molecular reality prior to 1900. J. Chem.

Edu. 51, 764–768.

Kloeden, P. E., Platen, E., 1992. Numerical solution of stochastic differential equations. Vol. 23 of Applications of Mathematics (New York). Springer-Verlag, Berlin.

Kopelman, R., 1986. Rate-processes on fractals – theory, simulations, and experiments. J. Stat. Phys. 42, 185–200.

Kopelman, R., 1988. Fractal reaction kinetics. Science 241, 1620–1626.

Kramers, H. A., 1940. Brownian motion in a field of force and the diffusion model of chemical reactions. Physica 7, 284–304.

Kuthan, H., 2001. Self-organisation and orderly processes by individual protein complexes in the bacterial cell. Prog. Biophy. Mol. Biol. 75, 1–17.

Luby-Phelps, K., Castle, P. E., Taylor, D. L., Lanni, F., 1987. Hindered diffusion of inert tracer particles in the cytoplasm of mouse 3T3 cells. Proc. Natl. Acad. Sci. USA 84, 4910–4913.

Majid, I., ben Avraham, D., Havlin, S., Stanley, H. E., 1984. Exact-enumeration approach to random walks on percolations clusters in two dimensions. Phys. Rev. B. 30, 1626–1628.

McAdams, H. H., Arkin, A., 1997. Stochastic mechanisms in gene expression. Proc. Natl. Acad. Sci. U. S. A. 94, 814–819.

Minton, A. P., 1993. Molecular crowding and molecular recognition. J. Mol. Recognit. 6, 211–214.

Minton, A. P., 1998. Molecular crowding: Analysis of effects of high concentrations of inert cosolutes on biochemical equilibria and rates in terms of volume exclusion. Methods Enzymol. 295, 127–149.

Morton-Firth, C. J., 1998. Stochastic simulation of cell signalling pathways. Ph.D. thesis, University of Cambridge.

Morton-Firth, C. J., Bray, D., 1998. Predicting temporal fluctuations in an intracellular signalling pathway. J. theor. Biol. 192, 117–128.

Qian, H., Elson, E. L., 2002. Single-molecule enzymology: stochastic Michaelis–Menten kinetics. Biophys. Chem. 101–102, 565–576.

Rao, C. V., Arkin, A. P., 2003. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. J. Chem. Phys. 118, 4999–5010.

Rathinam, M., Petzold, L. R., Cao, Y., Gillespie, D. T., 2003. Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method. J. Chem. Phys. 119, 12784–12794.

Resat, H., Wiley, H. S., Dixon, D. A., 2001. Probability-weighted dynamic Monte Carlo method for reaction kinetics simulations. J. Phys. Chem. B 105, 11026–11034.

Sano, Y., Shimada, T., Nakashima, H., Nicholson, R. H., Eliason, J. F., Kocarek, T. A., Ko, M. S. H., 2001. Random monoallelic expression of three genes clustered within 60 kb of mouse t complex genomic DNA. Genome Res. 11, 1833–1841.

Scalettar, B. A., Abney, J. R., Hackenbrock, C. R., 1991. Dynamics, structure, and functions are coupled in the mitrocondrial matrix. Proc. Natl. Acad. Sci. USA 88, 8057–8061.

Schnell, S., Maini, P. K., 2000. Enzyme kinetics at high enzyme concentration. Bull. Math. Biol. 62, 483–499.

Schnell, S., Maini, P. K., 2002. Enzyme kinetics far from the standard quasi-steady-state and equilibrium approximations. Math. Comput. Modelling 35, 137–144.

Schnell, S., Maini, P. K., 2003. A century of enzyme kinetics. Reliability of the K_M and $v_{\max}$ estimates. Comments Theor. Biol. 8, 169–187.

Schnell, S., Mendoza, C., 1997. Closed form solution for time-dependent enzyme kinetics. J. theor. Biol. 187, 207–212.

Schnell, S., Turner, T. E., 2004. Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws. Prog. Biophys. Mol. Biol. 85, 235–260.

Shimizu, T. S., Le Novere, N., Levin, M. D., Beavil, A. J., Sutton, B. J., Bray, D., 2000. Molecular model of a lattice of signalling proteins involved in bacterial chemotaxis. Nat. Cell Biol. 2, 792–796.

Srere, P., Jones, M. E., Mathews, C., 1989. Structural and organizational aspects of metabolic regulation. Alan R. Liss, New York.

Staff, P. J., 1970. A stochastic development of the reversible Michaelis–Menten mechanism. J. theor. Biol. 27, 221–232.

Tian, T. H., Burrage, K., 2004a. Binomial leap methods for simulating chemical kinetics. J. Chem. Phys., submitted.

Tian, T. H., Burrage, K., 2004b. Bistability and switching in the lysis/lysogeny genetic regulatory network of bacteriophage λ. J. theor. Biol. 227, 229–237.

Verkman, A. S., 2002. Solute and macromolecule diffusion in cellular aqueous compartments. Trends Biochem. Sci. 27, 27–33.